

# Upper Bounds on Optimization Time of Population-Based Evolutionary Algorithm on a Function with Fitness Plateaus Using Elitism Levels Traverse Mechanism

Aram Ter-Sarkisov      Stephen Marsland

Department of Computer Science  
Massey University, New Zealand  
`{a.ter-sarkisov,s.r.marsland}@massey.ac.nz`  
<http://www.massey.ac.nz/seat>

March 2, 2013

## Abstract

In this article a tool for the analysis of population-based EAs is used to derive asymptotic upper bounds on the optimization time of the algorithm solving Royal Roads problem, a test function with plateaus of fitness. In addition to this, limiting distribution of a certain subset of the population is approximated.

## 1 Introduction

Since early 90s a large amount of theoretical analysis of EA has evolved, both in convergence (rate of improvement) and runtime (expected optimization time). Most research in EA community so far has been focused on single-species (i.e.  $(1 + 1)$ EAs with some form of mutation). EAs with nontrivial size of the population and/or the recombination pool were analyzed in [HY02, HY03] (population-based algorithms) [CHS<sup>+</sup>09, CTCY11] (evolution of the is driven by the accumulation of a special type

of species, locally-optimal individuals (LOIs), rather than the probability of finding a better fit species), [Wit06, Wit08] (tree-based analysis), [TSM11b, TSM11a] (population-based EAs for Royal Roads (RR) test problem and 1-Bit-Swap operator).

One of the main questions that these findings tried to answer, was, if the effect of the population was positive. It was found, that, if measured in the number of function evaluations rather than generations, for easy problems with a large ( $O(n!)$ ) number of paths, population degrades performance. In [HY04]  $(n+n)$ EA with recombination, mutation and selection solves OneMax in  $O(n^2)$  if measured in the number of function evaluations vs  $\Theta(n \log n)$  by  $(1+1)$ EA. This means a population-based algorithm boosts performance only on parallel computers, i.e. if the fitness is computed on parallel computers simultaneously.

Later in [CHS<sup>+</sup>09] it was shown that when measured in the number of function evaluations, asymptotic order of population-based EAs compared to  $(1+1)$ EA are no worse only for population size  $O(1)$  (for problems like OneMax and Leading Ones). An advantage may come from the point of view of diversity, reducing the risk of bad initialization and the probability of finding a solution eventually (convergence probability).

Most of the research on population-based algorithms uses two tools to derive runtime:

1. Drift analysis. It was introduced in [HY01] and widely applied to both single-species (see i.e. [DJW10b, DJW11, Jag08] etc) as well as population-based EAs (i.e. [JDJW05, OHY08, NOW09]). Drift is a form of a martingale functions that is widely used in other areas of science (e.g. quantitative analysis, physics, etc) to bound the difference in change of a stochastic process (see e.g. [Haj82])
2. Family Trees. It was first introduced in EA community in [Wit06] and is based on bounding the distance of an offspring from the ancestor generated randomly at the start of an algorithm after a number of generations.

Most of the analysis (except [HY02, HY03, CHS<sup>+</sup>09, CTCY11, TSM11b, TSM11a]) was focused on either  $(\mu+)$  or  $(1+\lambda)$  algorithms. Naturally, the question arises, if the results for these two can be extended to a  $(\mu+\lambda)$ . The answer is probably no, as the result, in, e.g. [CHS<sup>+</sup>09] for  $(N+N)$ EA is (if measured in the number of function evaluations)  $O(nN \log N + n \log n)$ . It cannot be directly derived from combining the results for  $(\mu+1)$  in, e.g. [Wit06] and  $(1+\lambda)$  in [He10].

Another issue is that in most of these papers population or recombination pool were considered ‘monolithic’, i.e. species with different fitness were never considered separately. The disadvantage of this approach is obvious: the evolution of the population ‘on the whole’ is determined by the evolution of its subsets, and not just one type of species (e.g. currently best). Therefore runtime (optimization time) bounds can vary substantially by analyzing different subsets of the population.

In Section 2 we explain in details the motivation for the development of the new approach to the analysis of population-based algorithms. In Section 3 details of the analyzed algorithm and the problem are given. In Section 4 Elitism Levels Traverse Mechanism is presented. In Section 5 the birth-and-death Markov Chain for the runtime is derived. In Section 6 main results of this article are presented, i.e. the upper bounds on the runtime of the algorithm derived using the Elitism Levels Traverse Mechanism. In Section 7 we approximate the stationary distribution of super-elite species in the population. Section 8 concludes.

## 2 Complexity of Analysis and Approach

In Section 1 we outlined some of the drawbacks of the current approach to the analysis of  $(\mu + \lambda)$ EAs. Here we give a more detailed motivation for the development of an analytical tool suited for these algorithms.

It may seem strange that  $(\mu + \lambda)$  algorithms (with both  $\mu$  and  $\lambda > 1$ ), despite their widespread application in real life are less popular in theoretical EA community than  $(1 + 1)$ , some of them outlined in the previous Section. One of the main reason is the complexity of the structure of the population. Throughout the run of the algorithm both population and recombination pools consist of different types of species with different fitnesses. If a genetic operator that recombines information between parents is used, then pairs of parents have to be considered rather than single species. Quite obviously the structure and prevalence of certain types of species in the population affects greatly the dynamics of the evolutionary process.

More specifically, the complexity arising from the structure of the population results in the following issues:

1. Distribution of species of different types in the population. This is probably the most obvious consequence of using  $\mu > 1$ . Regardless of the problem con-

sidered (even with traps and/or local minima), species with different fitness do not have equal representation in the population. Among other things, it is important to distinguish species with currently best fitness (elite) and the rest, especially for functions with plateaus. To the best of my knowledge, this quite obvious and important fact has never attracted much interest in EA community, most likely due to the complexity of the combinatorial structures of these subsets of the population. Interesting though, it is well-known in biology, epidemiology and related areas.

2. Fitness-proportional selection. This selection is one of the most important features of EAs. Obviously, it defines the structure of recombination pool, and depends on the distribution of species in the population. If a certain type of species dominates the population, they may be underrepresented in the recombination pool due to low fitness. Thus, depending on the type of the problem (e.g. with multiple optima), it may greatly inhibit the evolutionary process. At the same time, currently best (or next-best) species, even though they need not be abundant in the population, have a high probability of expanding their presence in the recombination pool. This issue was touched upon in [CHS<sup>+</sup>09], where the number of elite species in the recombination pool was found to exceed that in the population by at least 25%. The selection process for any selection type was simplified to the proportion of LOI in the population and ignoring the upgrade of non-LOI to LOI. One can suspect that different selection functions lead to different structures of recombination pools and thus variance in the probability of generating better offsprings.
  
3. Pairing of parents. This point is valid for algorithms with recombination operators, e.g. crossover. The number of types of pairs greatly affects the probability of evolving higher-ranked offsprings and their structure. If the algorithm uses recombination pool size  $\lambda$ , there are  $\frac{\lambda}{2}$  pairs. Even for a simple case when the population has only two types of species,  $\alpha$  and  $\beta$ , there are three types of pairs:  $\langle \alpha, \alpha \rangle$ ,  $\langle \alpha, \beta \rangle$  (or  $\langle \beta, \alpha \rangle$ ) and  $\langle \beta, \beta \rangle$ , each with its own properties and evolution probabilities that can differ to a greater degree than an order of a constant. Even in this simple case there are, using multinomial coefficients  $\sum_{j=0}^{\frac{\lambda}{2}} \sum_{r=0}^{\frac{\lambda}{2}-j} \binom{\frac{\lambda}{2}}{j} \binom{\frac{\lambda}{2}-j}{r} = \sqrt{3}^{\lambda}$  possible combinations of pairs of parents in the recombination pool.

4. Exchange of genetic information between parents in the recombination pool. This point is also valid for algorithms with recombination operators. The majority of publications that consider populations with crossover (e.g. [OHY08]) only a simple lower bounds on the probability to improve fitness on each fitness level using crossover is considered, although the probability of success can vary substantially between different types of parents. Even if the analysis of different pairs of parents does not improve the asymptotic runtime result, it may shed light on some properties of the recombination process and the efficiency of EA, since it is hypothesized that EAs' performance is driven by recombination rather than mutation.
5. Distribution of offsprings and genetic (fitness) progress. Finally, little is known about the results of recombining genetic information, more specifically, the distribution of offsprings and the replacement of old population. Usually, as in [CHS<sup>+</sup>09] the best  $\mu$  species is selected from  $\mu + \lambda$  parents and offsprings. Depending on how 'progress' is defined: generation of one, at least one, or some specific number/proportion of better offsprings, it can lead to very different bounds on runtime.

Although many of these issues are known in theoretical EA community, they are usually overly simplified or avoided altogether. The main contribution of this article is the development of a new approach that allows to answer some of these questions.

### 3 Algorithm and the Test Function

The algorithm analyzed here uses the simplest variant of K-Bit-Swap operator introduced in [TSMH10] with  $K = 1$ . It is a recombination operator that transfers information between two parents and has some features of both uniform crossover and mutation.

The only test problem considered in this article is Royal Roads, a test function with plateaus of fitness. It was introduced and discussed in [MFH92, Mit96] and instances of it were analyzed in [WJ07, SW03]. A more general non-modular problem was also considered in [Rud98].

The feature that makes this type of problems quite hard for EA (originally they

were designed to test DA’s ability to recombine schemata, see [Mit96]) is the lack of guidance towards the global optimum, although unlike, e.g. TwoMax problem is does not have ‘traps’ that lead the algorithm into local optimum.

### 3.1 Algorithm

Below are Tables 1, 2, 3 with pseudocode for a  $(\mu + \lambda)$  Evolutionary algorithm using a variant of Tournament selection with a 1-Bit-Swap (1BS) genetic operator. For more details on 1BS see [TSMH10].

Table 1:  $(\mu + \lambda)$  Evolutionary Algorithm using 1-Bit-Swap

---

1	create $\mu$ starting species at random
2	$t = 0$
	<b>loop until the solution is found</b>
4	select using a variant of fitness-proportional Tournament selection $\frac{\lambda}{2}$ pairs of parents into the pool
5	swap bits in each pair using 1BS
6	keep the currently-elite species in the population, replace the rest with the pool, first with new currently-elite, then at random
7	$t = t + 1$
8	<b>end loop</b>

---

Table 2: Variant of Tournament Selection

---

1	$k = 0$
2	<b>loop over the size the pool</b>
3	select two species from the population at random
4	examine their fitness, the better one enters the pool
5	$k = k + 1$
6	<b>end loop</b>

---

### 3.2 Test problem

Our notation is somewhat different to that in [Mit96], but otherwise we follow the idea of the Royal Roads (RR) function quite closely. We consider a string of binary values length  $n$  which is divided into  $K$  disjoint subsets (bins) length  $M$  each s.t.  $n = KM$ . Although RR directly implies that the bins are consecutive, since we do

Table 3: 1-Bit-Swap Operator

---

1	$j = 0$
2	<b>loop over the number of pairs in the pool</b>
3	select a bit in the first parent uniformly at random
4	select a bit in the second parent uniformly at random
5	swap these bits between parents
5	$j = j + 1$
6	<b>end loop</b>

---

not use segment or uniform crossover, they need not be such. We denote  $x_{k,j}$  the value of the  $j$ 'th bit in the  $k$ 'th bin in string  $s$ . Therefore objective function of the string  $s$

$$f(s) = \sum_{k=0}^{K-1} c_k \prod_{j=0}^{M-1} x_{k,j} \quad (1)$$

w.l.o.g. we set  $c_k = M \forall k$ . Quite obviously the unique global maximum of fitness is then of course  $n$ . The most complicated feature of the problem is the fitness-proportional selection function that does not differentiate between species in the population if their fitness is the the same, i.e. species on the plateau of fitness. It is obvious though from Equation 1 that we need to somehow distinguish between strings of the same fitness  $kM$  but different distance to the next fitness level  $(k+1)M$ .

This problem is taken care of by introducing the auxiliary function, OneMax that simply sums the number of 1-bits in each bin. In essence, this allows to track the progress of currently-elite species to the next fitness plateau.

The second distinct feature of the algorithm analyzed here is that the offspring can either have -1, 0 or 1 more 1-bit anywhere after swapping. This means that it is impossible that 2 bins evolve simultaneously. In other words, bins in the population evolve in some arbitrary sequence (obviously there are  $K!$  such sequences). Therefore, we can ignore progress in other bins and only look at the evolution at some arbitrary  $k$ 'th bin that we refer to as active.

We are now ready to discuss the new approach we have developed.

## 4 Elitism Levels Traverse Mechanism

This new approach uses some previously-known ideas, such as artificial fitness levels (see [JW01]) and locally-optimal individuals (see [CHS<sup>+</sup>09]). This approach addresses some of the main problems revealed in Section 2 and suggests a different way to derive the upper bound on the mean first hitting time of the global solution. Its main features are:

1. Identification of the possible ways of adding elite species that can improve the fitness of the population. These ways are combined into the probability of a successful event
2. Elimination of any assumptions of the distribution of species in the population that make enables derivation of sharper runtime bounds
3. Instead of finding the probability of advancing a level of fitness, it works by finding the probability of advancing a level of elitism, e.g. adding an elite species up to a certain proportion  $\delta$  that ensures the  $1 - o(1)$  probability of evolution

The working of the Elitism Levels Traverse Mechanism can be illustrated by an example from immunology.

Suppose that there exists a population of species of size  $N$ , which is susceptible to  $M$  types of infection, which are mutually exclusive, i.e. a species cannot be infected by more than one infection at the same time. The size of each set of infected species cannot be larger than  $m_j$ . We denote an event that there are  $1 \leq r \leq m_j$  infected species of type  $1 \leq j \leq M$  that can infect exactly one healthy member of the population by  $E_j^*$ . Since the sets of infected species are mutually exclusive, by additivity the probability that a healthy species get any of the infections is obtained:

$$P\left(\bigcup_{j=1}^M E_j^*\right) = P\left(\bigcup_{j=1}^M \bigcup_{r=1}^{m_j} E_{jr}^*\right) = \sum_{j=1}^M \sum_{r=1}^{m_j} P(E_{jr}^*) = \sum_{j=1}^M P(E_j^*) = P(E^*)$$

This expression needs to be simplified for a number of reasons, e.g. the knowledge of  $m_j$ . Although one can find bounds on the partial sum of rows of Pascal triangle (since  $m_j$  is clearly less than  $N$ ), it is guaranteed to make the derivation messy. This is done by considering only one infected species of each type rather than  $r$  and the



event of infecting exactly one healthy species with any of  $M$  viruses is therefore  $E_j$ . This yields the lower bound on the total probability of adding exactly one infected offspring:

$$P(E^*) = \sum_{j=1}^M P(E_j^*) \geq \sum_{j=1}^M P(E_j) = P(E) \quad (2)$$

The proof of Equation 2 is in [TS12]. Obviously, this inequality becomes strict if there are at least two infected species with at least one type of infection

Expressing these ideas in the notation and language of EA, infected species are elite,  $M$  is the number of ways to generate elite offsprings,  $N = \frac{\lambda}{2}$ , the number of pairs in the recombination pool. Thus, the probability above given  $\alpha$  elite species and  $k$ 'th fitness level, becomes  $P(\alpha, k)$ . We need to find the time until there are  $\delta\mu$  elite species in the population, which ensures the  $1 - o(1)$  probability of advancing to the next fitness level. If there are at most  $n$  such levels, the upper bound on the runtime of the algorithm is

$$\mathbf{E}\tau < \sum_{k=1}^n \sum_{\alpha=1}^{\delta\mu} \frac{1}{P(E(\alpha, k))} \quad (3)$$

Derivation of the upper bound from Equation 3 is rather versatile. One needs to identify pairs of possible parents  $\langle p_1, p_2 \rangle$  such that there exists some strictly positive probability of swapping (or otherwise exchanging) bits  $\varphi(k)$  that, as a results, at most one new elite offspring evolves. Nevertheless, as explained in Section 3, the fitness-proportional selection function does not differentiate between species on the plateau of fitness.

As a result, Equation 3 cannot be directly applied to the Royal Roads test function, since elite species have different distance to the next fitness level, and some of the elite species can degrade due to random walk on the plateau. In the next subsection we give more details on the application of the Elitism Levels Traverse Mechanism to the Royal Roads test function.

## 4.1 Structure of the population (partition of the levels of elitism)

Both the Elitism Levels Traverse Mechanism and the evolutionary process of the algorithm described in the previous subsection depend predominantly on the structure

of the population, which is discussed here.

The run of the algorithm is divided into Phase 1 and 2. Phase 1 starts at the random initialization of the algorithm and ends when the first active bin is solved. Phase 2 starts with the second active bin and finishes when the whole problem is solved. The reasons for this partition are quite clear: until the first bin is solved, fitness function does not distinguish between species (they are all of the same fitness). After that the structure of population becomes more distinct, with elite and non-elite subsets. In both runs though we need to partition the elitism level, i.e. the level which includes only elite species by using the auxiliary function.

### Phase 1

To obtain probabilities for this Phase, we see that if measured by fitness function, all species (candidate parents) have the same fitness. Therefore in this Phase we can distinguish them only by the auxiliary function:

$\alpha^*$  : species with the highest auxiliary function (closest to the next fitness level)

$\beta^*$  : species with the next-best value of the auxiliary function

$\gamma^*$  : the rest of the population

In [CHS<sup>+</sup>09] it is suggested to ‘wait’ until population accumulates  $\frac{\mu}{2}$  currently-elite species and then find the probability that at least one of them evolves. We instead consider a certain number  $\delta^*\alpha$  (in Phase 1  $\mu = \alpha$  for the reasons pointed out above) of super-elite species such that the probability to evolve an offspring with higher auxiliary value is arbitrarily close to 1.

Also, due to the complexity of the structures arising in the dynamics of populations, we use a trivial lower bound on the number of  $\beta^*$  species, 1, and thus the number of  $\gamma^*$  is simply  $\mu - \alpha^* - 1$ . Since a the variant of Tournament selection is used, this means that if the fitness of the candidates is the same, the parent is selected randomly, we use the lower bound on selection probability. For example, the probability we use to select pair  $\langle \alpha^*, \beta^* \rangle$  is

$$\frac{\alpha^*}{\mu} \left( \frac{1}{2} + 1 \right) \cdot \frac{1}{\mu} \left( \frac{1}{2} + 1 \right) > \frac{\alpha^*}{4\mu^2}$$

Applying Elitism Levels Traverse Mechanism we find that there are four types of pairs that may add exactly one super-elite offspring to the population (since the

older super-elite species are saved due to elitism):

- $< \alpha^*, \beta^* >$  : need to select at least 1-bit in the active bin in the first parent  
and a 1-bit anywhere in the second.  
on the success probability
- $< \alpha^*, \gamma^* >$  : need to select a 1-bit in the active bin in the first parent  
and a 1-bit anywhere in the second
- $< \beta^*, \beta^* >$  : need to select a 0-bit in the active bin in either parent  
and a 1-bit anywhere in the the other
- $< \beta^*, \gamma^* >$  : need to select a 0-bit in the active bin of the first parent  
and 1-bit anywhere in the the other

Therefore, we need to bound the number of 1-bits ‘anywhere’ or in the active bin in the parents. We assume pessimistically that  $\alpha^*$  has at least three 1-bits in the active bin, therefore  $\beta^*$  has two and  $\gamma^*$  has one. This is to avoid division by 0. Also, when selecting 1-bit from ‘anywhere’ in the other parent, we consider the following: there are  $K - 1$  inactive bins, in which  $\beta^*$  parents have at least  $2(K - 1)$  1-bits and  $\gamma^*$  at least  $K - 1$  bits. Hence flipping probabilities  $\varphi_1 \dots \varphi_4$  are:

$$\begin{aligned}\varphi_1 &= \frac{j}{n} \cdot \frac{2(K-1) + j - 1}{n} \\ \varphi_2 &= \frac{j}{n} \cdot \frac{K-1 + j - 2}{n} \\ \varphi_3 &= 2 \cdot \frac{M-j+1}{n} \cdot \frac{2(K-1) + j - 1}{n} \\ \varphi_4 &= \frac{M-j+1}{n} \cdot \frac{K-1 + j - 2}{n}\end{aligned}$$

Similar logic applies to the loss probability, in Phase 1 it is just

- $< \alpha^*, \gamma^* >$  : need to select a 1-bit in the active bin in the first parent  
and a 0-bit anywhere in he second parent

which of course leads to the degrading of the super-elite species and reduction of their number by 1.

The runtime bounds for Phase 1 are derived in Subsection 6.2 and Appendix A.

## Phase 2

When the first bin is solved, population structure changes, since some species become better fit than the rest. From this point on, selection function distinguishes between elite  $\alpha$ , current highest fitness the rest of the population that we denote  $\eta$ . In addition to the four types of pairs mentioned above, we get two more:

- $< \alpha^*, \eta >$  : select a 0-bit in the active bin in the first parent and a 0-bit anywhere in the second parent
- $< \beta^*, \eta >$  : select a 0-bit in the active bin of the first parent and a 1-bit anywhere in the second parent

Later in Appendix C we show under what (quite general) conditions these two new types of pairs are lower-bounded by the ones defined in Phase 1. The complexity of this analysis comes from the fact that  $\eta$  species form the majority of population (since  $\alpha$  can be quite small), but we know nothing about their distribution (or make any specific assumptions thereof). Therefore, in the derivation of the lower bound in Appendix C, since all  $\eta$  species are worse than  $\alpha$ , the probability to select them is  $(1 - \frac{\alpha}{\mu})^2$  and the simple upper bound on the number of super-elite species is  $\alpha \geq \alpha^*$ . Then also a trivial lower bound on the probability of sampling a 1-bit anywhere in this type of species,  $\frac{K}{n}$ .

The probability to sample a 0-bit in  $\eta$  is upper-bounded by  $1 - \frac{K}{n}$ . These conditions seem quite loose, but nevertheless we are able to prove the lower bounds, albeit with some restrictions on  $\alpha$  proportion of the population.

Additionally, in Phase 2 by summing of  $k$ , the number of filled bins we account for the fact that we can sample an increasing number of 1-bits. Combining it with the same assumption of the number of 1-bits in these bins, we obtain the number of 0-bits in  $\gamma^*$  :  $n - (K - 1) - k(M - 1)$ .

In a similar way as in Phase 1, one more type of pairs that enables loss of a super-elite species is

- $< \alpha^*, \eta >$  : select a 0-bit in the active bin of the first parent and a 1-bit anywhere in the second parent

The details of the derivation for Phase 2 are in Subsection 6.2 and Appendix B.

Some features are common to both phases: monotonicity of probabilities (see Subsection 6.1) that enables us to lower-bound probabilities of adding a super-elite offspring to the population and an assumption that an  $\alpha$  species cannot degrade, e.g. by replacing a 1-bit with 0-bit in the bins that had already evolved.

One important simplification we use here is that the auxiliary function of the active bin is non-decreasing, which amounts to saying that the number of super-elite species cannot drop to 0, which is especially important in the construction of the Markov chain in Section 5. This can be justified by considering the following case: if the auxiliary function degrades, the number of  $\beta^*$  species is at least  $\delta^*\alpha$ , which by assumption means that the  $\alpha^*$  offspring is regenerated with probability  $1 - o(1)$ . Similar cases are well-known in areas such as immunology and ecology.

## 5 Markov chain analysis of the algorithm

We construct on each bit in the active bin, i.e. every level of auxiliary function a (auxiliary) birth-and-death Markov chain (MC) with 1 absorbing state ( $\delta^*\alpha$ ). The  $\delta^*\alpha \times \delta^*\alpha$  transition matrix for this MC is

$$\mathbf{P} = \begin{pmatrix} r_{1,1} & p_{1,2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ q_{2,1} & r_{2,2} & p_{2,3} & 0 & \cdots & 0 & 0 & 0 \\ 0 & q_{3,2} & r_{3,3} & p_{3,4} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q_{\delta^*\alpha-1, \delta^*\alpha-2} & r_{\delta^*\alpha-1, \delta^*\alpha-1} & p_{\delta^*\alpha-1, \delta^*\alpha} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

The dimensionality follows from the pessimistic assumption that each auxiliary level starts with only one super-elite parent. The expected first hitting time of the absorbing state from any state in the MC is expressed as a recurrent (difference) equation:

$$m_{\alpha^*, \delta^*\alpha} = 1 + q_{\alpha^*, \alpha^*-1} m_{\alpha^*-1, \delta^*\alpha} + r_{\alpha^*, \alpha^*} m_{\alpha^*, \delta^*\alpha} + p_{\alpha^*, \alpha^*+1} m_{\alpha^*+1, \delta^*\alpha} \quad (4)$$

due to the assumption this becomes

$$m_{1, \delta^*\alpha} = 1 + r_{1,1} m_{1, \delta^*\alpha} + p_{1,2} m_{2, \delta^*\alpha} \quad (5)$$

with a boundary condition  $m_{\delta^*\alpha, \delta^*\alpha} = 0$ . We define a new recurrence expression:

$$M_{\alpha^*} = m_{\alpha^*, \delta^*\alpha} - m_{\alpha^*+1, \delta^*\alpha}$$

trivially this quantity is nonnegative and its telescoping sum is

$$\sum_{\alpha^*=1}^{\delta^*\alpha-1} M_{\alpha^*} = m_{1,\delta^*\alpha}$$

We rewrite Equation 4 (since  $p + q + r = 1$ ):

$$\begin{aligned} (p_{\alpha^*,\alpha^*+1} + q_{\alpha^*,\alpha^*-1})m_{\alpha^*,\delta^*\alpha} &= 1 + q_{\alpha^*,\alpha^*-1}m_{\alpha^*-1,\delta^*\alpha} + p_{\alpha^*,\alpha^*+1}m_{\alpha^*+1,\delta^*\alpha} \\ p_{\alpha^*,\alpha^*+1}(m_{\alpha^*,\delta^*\alpha} - m_{\alpha^*+1,\delta^*\alpha}) &= 1 + q_{\alpha^*,\alpha^*-1}(m_{\alpha^*-1,\delta^*\alpha} - m_{\alpha^*,\delta^*\alpha}) \\ p_{\alpha^*,\alpha^*+1}M_{\alpha^*} &= 1 + q_{\alpha^*,\alpha^*-1}M_{\alpha^*-1} \\ M_{\alpha^*} &= \frac{1}{p_{\alpha^*,\alpha^*+1}} + \frac{q_{\alpha^*,\alpha^*-1}}{p_{\alpha^*,\alpha^*+1}}M_{\alpha^*-1} \end{aligned} \quad (6)$$

We also rewrite Equation 5 as:

$$m_{1,\delta^*\alpha} = \frac{1}{p_{1,2}} + m_{2,\delta^*\alpha}$$

and, therefore,

$$M_1 = \frac{1}{p_{1,2}}$$

Solving Equation 6 recurrently, we get the expression for the general term  $M_{\alpha^*}$ :

$$\begin{aligned} M_{\alpha^*} &= \frac{1}{p_{\alpha^*,\alpha^*+1}} \left( 1 + \frac{q_{\alpha^*,\alpha^*-1}}{p_{\alpha^*-1,\alpha^*}} + \dots + \frac{q_{\alpha^*,\alpha^*-1} \cdot q_{\alpha^*-1,\alpha^*-2} \cdot \dots \cdot q_{2,1}}{p_{\alpha^*-1,\alpha^*} \cdot p_{\alpha^*-2,\alpha^*-1} \cdot p_{1,2}} \right) \\ &= \frac{1}{p_{\alpha^*,\alpha^*+1}} \left( 1 + \sum_{m=2}^{\alpha^*} \prod_{l=0}^{\alpha^*-m} \frac{q_{\alpha^*-l,\alpha^*-l-1}}{p_{\alpha^*-l-1,\alpha^*-l}} \right) \end{aligned} \quad (7)$$

Summing on  $\alpha^*$ , the LHS is simply the desired quantity,  $m_{1,\delta^*\alpha}$ :

$$m_{1,\delta^*\alpha} = \sum_{\alpha^*=1}^{\delta^*\alpha-1} \frac{1}{p_{\alpha^*,\alpha^*+1}} + \sum_{\alpha^*=1}^{\delta^*\alpha-1} \frac{1}{p_{\alpha^*,\alpha^*+1}} \sum_{m=2}^{\alpha^*} \prod_{l=0}^{\alpha^*-m} \frac{q_{\alpha^*-l,\alpha^*-l-1}}{p_{\alpha^*-l-1,\alpha^*-l}} \quad (8)$$

Now the RHS has to be simplified to be solved. The numerator of each fraction is a product of probabilities, so each of them is upper-bounded by  $q_{\alpha^*,\alpha^*-1}$ . The denominator can be simplified in the following way: if the probability to increase the number of super-elite species grows (we prove this in Subsection 6.1) for any  $\alpha^* \geq 1$ , we can use the sequence of inequalities  $p_{\delta^*\alpha-1,\delta^*\alpha} \geq p_{\delta^*\alpha-2,\delta^*\alpha-1} \geq \dots \geq p_{1,2} \geq p_{1,2}^{\delta^*\alpha}$ .

Thus the upper bound bound on the first hitting time of  $\delta^*\alpha$  super-elite species in the population is

$$m_{1,\delta^*\alpha} \leq \sum_{\alpha^*=1}^{\delta^*\alpha-1} \frac{1}{p_{\alpha^*,\alpha^*+1}} + \frac{\delta^*\alpha}{p_{1,2}^{\delta^*\alpha}} \sum_{\alpha^*=2}^{\delta^*\alpha-1} q_{\alpha^*,\alpha^*-1} \quad (9)$$

From this we will derive the upper bound on the mean first hitting time of the algorithm (each  $m_{1,\delta^*\alpha}$  is also a function of the  $j$ , the number of bits set to 1 and  $k$ , the current plateau of fitness:

$$\begin{aligned} \mathbf{E}_{\mathcal{T}_{(\mu+\lambda)EA_{1BS}}} &= \sum_{k=0}^{K-1} \sum_{j=3}^{M-1} m_{1,\delta^*\alpha}(j, k) = \sum_{k=0}^{K-1} \sum_{j=3}^{M-1} \sum_{\alpha^*=1}^{\delta^*\alpha-1} \frac{1}{p_{\alpha^*,\alpha^*+1}} \\ &\quad + \sum_{k=0}^{K-1} \sum_{j=3}^{M-1} \frac{\delta^*\alpha}{p_{1,2}^{\delta^*\alpha}} \sum_{\alpha^*=2}^{\delta^*\alpha-1} q_{\alpha^*,\alpha^*-1} \end{aligned} \quad (10)$$

For simplicity we refer to the first term as the **first expression** and the second term as the **second expression**.

## 6 Upper Bounds on the Optimization Time on the Royal Roads Test Function

We start by showing the monotonicity of success probabilities that greatly simplify the derivation of the asymptotic upper bound.

### 6.1 Monotonicity of Selection probabilities

Earlier we used monotonicity to upper-bound the expression in the MC. Here we prove this statement. We start by simplifying all swap probabilities  $\varphi_1 \dots \varphi_4 = O(1)$ . We need to show

$$\frac{p_{\alpha^*,\alpha^*+1}}{p_{\alpha^*-1,\alpha^*}} \geq 1$$

The numerator and denominator for all four types of probabilities are (since binomial coefficient  $\binom{\lambda}{1}$  cancels out):

$$\begin{aligned} p_{\alpha^*, \alpha^*+1} &= \frac{2\alpha^*}{4\mu^2} + \frac{2\alpha^*(\mu - \alpha^* - 1)}{4\mu^2} + \frac{1}{4\mu^2} + \frac{2(\mu - \alpha^* - 1)}{4\mu^2} \\ p_{\alpha^*-1, \alpha^*} &= \frac{2(\alpha^* - 1)}{4\mu^2} + \frac{2(\alpha^* - 1)(\mu - \alpha^*)}{4\mu^2} + \frac{1}{4\mu^2} + \frac{2(\mu - \alpha^*)}{4\mu^2} \\ \frac{p_{\alpha^*, \alpha^*+1}}{p_{\alpha^*-1, \alpha^*}} &= \frac{2\mu - 2\alpha^* + 2\mu\alpha^* - 2\alpha^{*2} - 1}{2\alpha^* + 2\mu\alpha^* - 2\alpha^{*2}} - 1 \geq 1 \text{ if } \alpha^* \leq \frac{\mu}{2} \end{aligned}$$

We show in Section 6.3 that this inequality holds with high probability, i.e. we need only a small number of super-elite species ( $\delta^*\alpha$  is independent of  $\mu$  for  $\lambda = \mu$ ) to produce a string with a higher value of auxiliary function.

## 6.2 Upper bounds for $(\mu + \lambda)\mathbf{EA}_{1BS}$

### Phase 1

In Phase 1 we are concerned only with the first bin, so we sum over  $\alpha^*$  and  $j$ , the number of bits in the first bin.

To use Elitism Levels Traverse Mechanism, we need to identify all pairs of parents that may add exactly one super-elite string to the population. These are (as shown in Subsection 4.1)  $\langle \alpha^*, \beta^* \rangle, \langle \alpha^*, \gamma^* \rangle, \langle \beta^*, \beta^* \rangle, \langle \beta^*, \gamma^* \rangle$ . As we are primarily concerned with the evolution of super-elite species, we always use the trivial lower bound of  $\beta^* \geq 1$ . Therefore the number of  $\gamma^*$  species in  $\mu - \alpha^* - 1$ .

The probability of success is

$$p_{\alpha^*, \alpha^*+1} = \binom{\frac{\lambda}{2}}{1} \left( \frac{2\alpha^*}{4\mu^2} \varphi_1 + \frac{2\alpha^*}{2\mu} \left( \frac{\mu - 1 - \alpha^*}{2\mu} \right) \varphi_2 + \frac{1}{4\mu^2} \varphi_3 + \frac{2}{2\mu} \left( \frac{\mu - 1 - \alpha^*}{2\mu} \right) \varphi_4 \right)$$

with the same notation for the probabilities of swapping bits as in the proof of monotonicity. The rest of the derivation is in Appendix A. The upper bound on the runtime of Phase 1 is

$$\mathbf{ET}_1 \leq \frac{4\mu n^2 \log\left(\frac{\delta^*\alpha\mu}{\mu - \delta^*\alpha}\right) \log(M+1)}{\lambda K} + \frac{(\delta^*\alpha(M-4))^2}{1 - \frac{\lambda\delta^*\alpha M(2K+M)}{\mu n^2}} \quad (11)$$



### Phase 2

In this phase we are concerned with the remaining  $K - 1$  bins in the string, so in addition to  $\alpha^*$  and  $j$  we also consider  $k$ 'th active bin. In addition to the four pairs we had in Phase 1, we also get two more:  $\langle \alpha^*, \eta \rangle, \langle \beta^*, \eta \rangle$ .  $\eta$  is any species that is not in the subset of currently-elite. Since all the expressions are already quite messy, we try to simplify further. In Appendix C we prove the lower bound for these pairs and show the conditions under which they holds.

The expected first hitting time for Phase 2 is

$$\begin{aligned} \mathbf{ET}_2 \leq & \frac{8\mu n^2 \log \delta^* \alpha \log K \log(M+1)}{\lambda M} \\ & + \frac{(M-4)(K-2)^2 \lambda (M^2 - M - 6)(3\mu - 2\delta^* \alpha - 1)(\delta^* \alpha)^3 (2n - K - KM + M - 3)}{\left(1 - \frac{\lambda \delta^* \alpha M (2K+M)}{\mu n^2}\right) 96\mu^2 n^2} \end{aligned} \quad (12)$$

### Combined runtime of the algorithm

Combining the expressions in Equations 11 and 12, the upper bound on the expected first hitting time for the whole algorithm is obtained:

$$\begin{aligned} \mathbf{ET} = \mathbf{ET}_1 + \mathbf{ET}_2 \leq & \frac{4\mu n^2 \log \left(\frac{\delta^* \alpha \mu}{\mu - \delta^* \alpha}\right) \log(M+1)}{\lambda K} + \frac{(\delta^* \alpha (M-4))^2}{1 - \frac{\lambda \delta^* \alpha M (2K+M)}{\mu n^2}} \\ & + \frac{8\mu n^2 \log \delta^* \alpha \log K \log(M+1)}{\lambda M} \\ & + \frac{(M-4)(K-2)^2 \lambda (M^2 - M - 6)(3\mu - 2\delta^* \alpha - 1)(\delta^* \alpha)^3}{1 - \frac{\lambda \delta^* \alpha M (2K+M)}{\mu n^2}} \\ & \cdot \frac{2n - K - KM + M - 3}{96\mu^2 n^2} \end{aligned} \quad (13)$$

Equation 13 looks quite cumbersome, but it can be reduced by setting  $K = M = \sqrt{n}$  and taking  $\delta^* \alpha = O(1)$  as for OneMax in [TS12]:

$$\begin{aligned} \mathbf{ET} & \leq \frac{4\mu n^{\frac{3}{2}} \log n}{\lambda} \cdot O(1) + n \cdot O(1) + \frac{8\mu n^{\frac{3}{2}} \log^2 n}{\lambda} \cdot O(1) + \frac{\lambda n}{\mu} \cdot O(1) \\ & = O\left(\frac{\mu n^{\frac{3}{2}} \log^2 n}{\lambda}\right) \end{aligned} \quad (14)$$

or, measured in the number of function evaluations,

$$\mathbf{E}\tau = O(\mu n^{\frac{3}{2}} \log^2 n) \quad (15)$$

Comparing this result to similar in literature, including  $2^{\sqrt{n}} \log n$  in [Mit96] and  $O(n^2)$  in [SW03], it is easy to notice that  $(\mu + \lambda)\text{EA}_{1BS}$  outperforms other algorithms. Another interesting point is that the asymptotic runtime for Phase 2 is larger than the one for Phase 1 only by  $O(\log n)$ .

### 6.3 Proof of the upper bound of the number of super-elite species (the lower bound of the probability of evolution)

In the previous subsection the derivation of the runtime was based on the assumption that  $\delta^* \alpha = O(1)$ , i.e. some constant  $c$  that does not depend on  $\alpha$  or  $\mu$ . Here the bound on the probability to evolve a higher-ranked offspring is derived. The attention is restricted to Phase 1 only. Analysis on Phase 2 is similar.

We use the Law of total probability on the probability of failure (F), i.e. the probability that a species with higher auxiliary value does not evolve. We have three types of pairs that can evolve:  $\langle \alpha^*, \alpha^* \rangle$ ,  $\langle \alpha^*, \beta^* \rangle$ ,  $\langle \alpha^*, \gamma^* \rangle$ . We define event A that none of the three pairs get selected into the pool and event B that at least one of any pair gets selected. Since obviously  $P(F|A) = 1$ ,

$$P(F) = P(F|A)P(A) + P(F|B)P(B) = P(A) + P(F|B)P(B)$$

If there are  $c$  super-elite species in the population, the probability not to select any  $\langle \alpha^*, \alpha^* \rangle$  pairs is simply  $(1 - \frac{c^2}{4\mu^2})^{\frac{\lambda}{2}}$ . In a similar way, the probability no to select any of the other two types of pairs is, respectively,  $(1 - \frac{c}{4\mu^2})^{\frac{\lambda}{2}}$  and  $(1 - \frac{\mu - c - 1}{4\mu^2})^{\frac{\lambda}{2}}$  using the trivial lower bound on the number of  $\beta^*$  parents ( $> 1$ ). The product of these probabilities is

$$\begin{aligned} P(A) &\leq \left(1 - \frac{c^2}{4\mu^2}\right)^{\frac{\lambda}{2}} \left(1 - \frac{c}{4\mu^2}\right)^{\frac{\lambda}{2}} \left(1 - \frac{c(\mu - c - 1)}{4\mu^2}\right)^{\frac{\lambda}{2}} \\ &\leq e^{-\frac{c^2\lambda}{8\mu^2}} e^{-\frac{c\lambda}{8\mu^2}} e^{-\frac{c\lambda(\mu - c - 1)}{8\mu^2}} = e^{-\frac{\lambda\mu c}{8\mu^2}} \end{aligned}$$

For  $\lambda = \mu$ , which is a usual choice, the probability of this event becomes upper bounded:

$$P(A) \leq e^{-\frac{c}{8}}$$

the number of pairs of each type in the recombination pool is upper-bounded by (respectively)  $m_1, m_2, m_3$ , so the second part of the expression is the probability to select  $p$  pairs into the recombination pool and flip the bits unsuccessfully, which is denoted by  $\varphi'_1, \varphi'_2, \varphi'_3$ .

$$\begin{aligned} P(F|B)P(B) &= \sum_{p=1}^{m_1} \binom{\frac{\lambda}{2}}{p} \left( \frac{c^2 \varphi'_1}{4\mu^2} \right)^p \left( 1 - \frac{c^2}{4\mu^2} \right)^{\frac{\lambda}{2}-p} \sum_{p=1}^{m_2} \binom{\frac{\lambda}{2}}{p} \left( \frac{c \varphi'_2}{4\mu^2} \right)^p \left( 1 - \frac{c}{4\mu^2} \right)^{\frac{\lambda}{2}-p} \\ &\quad \cdot \sum_{p=1}^{m_3} \binom{\frac{\lambda}{2}}{p} \left( \frac{c(\mu - c - 1) \varphi'_3}{4\mu^2} \right)^p \left( 1 - \frac{c(\mu - c - 1)}{4\mu^2} \right)^{\frac{\lambda}{2}-p} \\ &\leq \left( 1 - \frac{c^2}{4\mu^2} \right)^{\frac{\lambda}{2}} \left( 1 - \frac{c}{4\mu^2} \right)^{\frac{\lambda}{2}} \left( 1 - \frac{c(\mu - c - 1)}{4\mu^2} \right)^{\frac{\lambda}{2}} \left( \sum_{p=0}^{m_1} \frac{(\frac{\lambda}{2})^p}{p!} \left( \frac{c^2 \varphi'_1}{4\mu^2 - c^2} \right)^p - 1 \right) \\ &\quad \cdot \left( \sum_{p=0}^{m_2} \frac{(\frac{\lambda}{2})^p}{p!} \left( \frac{c \varphi'_2}{4\mu^2 - c} \right)^p - 1 \right) \left( \sum_{p=0}^{m_3} \frac{(\frac{\lambda}{2})^p}{p!} \left( \frac{c(\mu - c - 1) \varphi'_3}{4\mu^2 - c} \right)^p - 1 \right) \\ &\leq e^{-\frac{c}{8}} \left( e^{\frac{\lambda c^2 \varphi'_1}{2(4\mu^2 - c^2)}} - 1 \right) \left( e^{\frac{\lambda c \varphi'_2}{2(4\mu^2 - c)}} - 1 \right) \left( e^{\frac{\lambda c(\mu - c - 1) \varphi'_3}{2(4\mu^2 - c(\mu - c - 1))}} - 1 \right) \end{aligned}$$

The front term is  $O(1)$ . We take  $\max\{\varphi'_1, \varphi'_2, \varphi'_3\} = \varphi' \leq 1 - O(\frac{1}{n})$  and obtain the upper bound for the exponential term in the first bracket  $e^{\frac{\lambda c^2}{2(4\mu^2 - c^2)} - \frac{\lambda c^2}{2n(4\mu^2 - c^2)}}$  and  $e^{\frac{\lambda c}{2(4\mu^2 - c)} - \frac{\lambda c}{2n(4\mu^2 - c)}}$  in the second bracket. Asymptotically for  $\mu, n \rightarrow \infty$  and  $\mu = \lambda$  both of these terms converge to 1, so expressions in both brackets are  $o(1)$ . Following the same ideas, the exponential term in the last bracket is  $O(1)$ , so the whole expression is

$$P(F|B)P(B) = O(1) \cdot O(1) \cdot o(1) \cdot o(1) = o(1)$$

Combining the results above, we obtain the probability of failing evolution given a constant number of super-elite species in the population is

$$P(F) \leq e^{-\frac{c}{8}} + o(1) \tag{16}$$

And, therefore with probability of at least  $1 - e^{-\frac{c}{8}} + o(1)$  the population evolves by generating a higher-ranked offspring. For  $c = 1$  this probability is roughly 0.1175

and for  $c = 6 \approx 0.5276$ .

In this Section we have looked at how the evolution of the super-elite species affects the evolution of the population. What we want to do next is look at the evolution of these species, i.e. determine if there is any limiting (i.e. if the algorithm is run for a very long time) distribution to which their number converges.

## 7 Approximation of quasi-stationary distribution of the number of super-elite species

In certain areas of sciences, such as epidemiology and the study of computer viruses, Markov chain models are being widely applied to understand the propagation and extinction of processes, e.g. spread of viral diseases (see [Nas99]), the distribution of uninfected computers in a computer network (see [WM04]), etc. One of the best known models in this area is Susceptible-Infected-Susceptible (SIS, see [Nas99, WM04]), which roughly corresponds to the birth-and-death MC we have constructed in Section 5. To the best of our knowledge this is the first application of this approach in EA community, although statistical distribution of the first hitting times was analyzed in [GKS99].

In this Section we find an approximation to the limiting distribution of the number of super-elite species in the population. Derivation of the upper bound were shown to hold with high probability if  $\delta^*\alpha = O(1)$ . For the purpose of this Section we relax this requirement and obtain an asymptotic approximation of the stationary distribution of the MC, i.e. with  $\delta^*\alpha \rightarrow \infty$  and also  $\delta^*\alpha = o(\mu)$ .

Markov chain is the same as in Section 5, which is aperiodic, irreducible and time-homogeneous. Since the presence of the absorbing state makes the stationary distribution trivial with the full mass of probability set to state  $\delta^*\alpha$ , we transform the MC adding the probability of moving from state  $\delta^*\alpha$  back to state 1 equal to 1, hence this distribution is quasi-stationary. In the light of the set-up of the whole model this makes sense, since after the improvement in the auxiliary function the new number of super-elite species reduces back to 1.

We now also assume it is reversible (otherwise we need to derive a set of recurrent equations similar to those in Section 5). Stationary distribution of an MC is defined as the limiting proportion of time spent by the stochastic process  $X_t$  in a

state  $s_k$ :

$$\pi_k = \lim_{t \rightarrow \infty} P(X_t = s_k)$$

From this, using the set of detailed balance equations we can quite easily derive the expressions for the stationary distribution:

$$\begin{aligned} p_{1,2}\pi_1 &= q_{2,1}\pi_2 \\ p_{2,3}\pi_2 &= q_{3,2}\pi_3 \\ &\dots\dots\dots \\ p_{\delta^*\alpha-2,\delta^*\alpha-1}\pi_{\delta^*\alpha-2} &= q_{\delta^*\alpha-1,\delta^*\alpha-2}\pi_{\delta^*\alpha-1} \\ \pi_2 &= \frac{p_{1,2}}{q_{2,1}}\pi_1 \\ \pi_3 &= \frac{p_{1,2}p_{2,3}}{q_{2,1}q_{3,2}}\pi_1 \\ &\dots\dots\dots \\ \pi_{\alpha^*} &= \frac{p_{1,2}p_{2,3}\dots p_{\alpha^*-1,\alpha^*}}{q_{2,1}q_{3,2}\dots q_{\alpha^*,\alpha^*-1}}\pi_1 = \pi_1 \prod_{l=1}^{\alpha^*} \frac{p_{l-1,l}}{q_{l,l-1}} \end{aligned}$$

and, since  $\pi_{\alpha^*}$  is a probability distribution:

$$\sum_{\alpha^*=1}^{\delta^*\alpha} \pi_{\alpha^*} = 1$$

which enables us to find the expression for  $\pi_1$ :

$$\pi_1 = \frac{1}{1 + \sum_{m=2}^{\delta^*\alpha-1} \prod_{l=2}^m \frac{p_{l-1,l}}{q_{l,l-1}}}$$

The rest of the article is dedicated to finding asymptotic approximations of these stationary distributions and discussion of what they tell us about convergence of EA. We consider here two cases, one being a slow rate of progress, the other a fast rate of progress. All derivations are for Phase 1. Findings for Phase 2 are very similar.

## 7.1 Slow progress rate (Poisson approximation)

Derivation of  $\pi_{\alpha^*}$  depends on the ratio  $\frac{p_{l-1,l}}{q_{l,l-1}}$ . This ratio can be written as a product

$$\frac{p_{l-1,l}}{q_{l,l-1}} = \frac{r(l)\theta_1}{s(l)\theta_2} \approx \frac{\rho}{l}$$

where  $\theta_1, \theta_2$  do not depend on  $l$ . If the approximation above holds, and  $\frac{\theta_1}{\theta_2} = \rho$  is not very large, the following limiting distribution of super-elite species can be derived.

$$\pi_1 = \pi_1 = \frac{1}{1 + \sum_{m=2}^{\delta^* \alpha} \prod_{l=2}^m \frac{p_{l-1,l}}{q_{l,l-1}}} \approx \frac{1}{1 + \sum_{l=1}^{\delta^* \alpha} \frac{\rho^l}{l!}} = c^* e^{-\rho}$$

for some constant  $c^*$ . The ratio in the denominator accounts for quasi-stationarity:  $p_{\delta^* \alpha, 1} = 1$ . Therefore, the stationary distribution for  $\alpha^*$  super-elite species is

$$\pi_{\alpha^*} = c^* e^{-\rho} \frac{\rho^{\alpha^*}}{\alpha^*!}$$

which is a form of truncated Poisson distribution with removed origin (0) and the upper tail ( $\delta^* \alpha, \delta^* \alpha + 1, \dots$ ). If we allow  $\delta^* \rightarrow \infty$ ,  $c^*$  can be found:

$$c^* = \frac{1}{\sum_{\alpha^*=1}^{\infty} \pi_{\alpha^*}} = \frac{1}{1 - e^{-\rho}}$$

## 7.2 Fast progress rate(Normal approximation)

Here we consider the limiting distribution of super-elite species as  $\rho \rightarrow \infty$ . We use  $A^*$  as a random variable for super-elite species:

$$\pi_{\alpha^*} = \mathbf{P}(A^* = \alpha^*) = \frac{c^* e^{-\rho} \rho^{\alpha^*}}{\alpha^*!} = \frac{\rho^{\alpha^*}}{\alpha^*!} \pi_1$$

which, as mentioned before, is a form of truncated Poisson distribution with removed origin (for comparison see [Nas96], Section 3). We use characteristic function to derive its expectation and variance:

$$\phi_{A^*}(t) = \mathbf{E} e^{itA^*} = \sum_{\alpha^*=1}^{\delta^* \alpha} e^{it\alpha^*} \pi_{\alpha^*} = \pi_1 \sum_{\alpha^*=1}^{\delta^* \alpha} e^{it\alpha^*} \frac{\rho^{\alpha^*}}{\alpha^*!} = \pi_1 e^{\rho e^{it}}$$

For  $\delta^* \alpha \rightarrow \infty$ . Standardizing constant is  $c^*$  (to account for removed origin). By taking a derivative w.r.t  $t$  and setting  $t = 0$ :

$$\mathbf{E} A^* = \frac{1}{i} \phi'_{A^*}(0) = \frac{d \mathbf{E} e^{itA^*}}{i dt} \Big|_{t=0} = \pi_1 \rho e^{\rho e^{it}} e^{it} \Big|_{t=0} = \pi_1 e^{\rho} \rho = c^* \rho$$

In a similar way we find asymptotic expression for the variance,  $\mathbf{Var} A^*$ :

$$\mathbf{Var} A^* = \mathbf{E} A^{*2} - (\mathbf{E} A^*)^2 = -\phi''(0) + (\phi'(0))^2$$

This uses the fact that  $i^2 = -1$  and  $\frac{1}{i} = -i$ . Therefore,

$$\begin{aligned}\phi''(t) &= i\pi_1\rho(ie^{it+\rho e^{it}} + i\rho e^{2it+\rho e^{it}}) = -\pi_1\rho(e^{it+\rho e^{it}} + \rho e^{2it+\rho e^{it}}) \\ \phi''(0) &= -\pi_1\rho(\rho + 1)e^\rho, (\phi'(0))^2 = -\pi_1^2\rho^2e^{2\rho}\end{aligned}$$

Hence,

$$\mathbf{Var}A^* = \pi_1\rho(\rho + 1)e^\rho - \pi_1^2\rho^2e^{2\rho} = c^*\rho(1 + \rho(1 - c^*))$$

Since we now know both expectation and variance of this random variable, we can find the limiting distribution of standardized  $A^*$ :

$$A' = \frac{A^* - \mathbf{E}A^*}{\sqrt{\mathbf{Var}A^*}}$$

Using the the sum of Poisson random variables we get  $\mathbf{E}A' = 0$ ,  $\mathbf{Var}A' = 1$ . Therefore:

$$\begin{aligned}A' &= \frac{\sum_l A_l^* - (\mathbf{E}A_1^* + \mathbf{E}A_1^* + \dots \mathbf{E}A_\rho^*)}{\sqrt{\rho\mathbf{Var}A_l^*}} = \frac{A_1^* - \mathbf{E}A_1^*}{\sqrt{\rho\mathbf{Var}A_l^*}} + \frac{A_2^* - \mathbf{E}A_2^*}{\sqrt{\rho\mathbf{Var}A_l^*}} + \dots \frac{A_\rho^* - \mathbf{E}A_\rho^*}{\sqrt{\rho\mathbf{Var}A_l^*}} \\ &= A'_1 + A'_2 + \dots A'_\rho\end{aligned}$$

where all  $A'_l$  are iid. The characteristic function of  $A'$  is

$$\begin{aligned}\phi_{A'}(t) &= \mathbf{E}e^{itA'} = \mathbf{E}e^{it\sum_l A'_l} = \mathbf{E}\prod_{l=1}^\rho e^{itA'_l} = (\mathbf{E}e^{itA'_l})^\rho = \left(\mathbf{E}\left(1 + itA'_l + \left(\frac{itA'_l}{2!}\right)^2 + O\left(\frac{1}{\rho^2}\right)\right)\right)^\rho \\ &= \left(1 - \frac{t^2}{2\rho}\right)^\rho \rightarrow e^{-\frac{t^2}{2}}\end{aligned}$$

which is a characteristic function of standard Normal distribution with parameters 0 and 1. The derivation was due to Taylor series expansion of the exponential function around 0 (since the function is dominated by  $\rho$ ). Since  $A'_l$  are all identically distributed, their expectation is 0. Additionally,

$$\begin{aligned}\mathbf{E}A'_l &< \mathbf{E}A^* < \infty \\ \mathbf{E}(A'_l)^2 &= \frac{\mathbf{E}(A_l^* - \mathbf{E}A_l^*)^2}{\rho\mathbf{Var}A_l^*} = \frac{\mathbf{Var}A_l^*}{\rho\mathbf{Var}A_l^*} = \frac{1}{\rho} < \infty\end{aligned}$$

So the conditions for the Central Limit Theorem are fulfilled, and super-elite species converge to Normal distribution if the progress rate is high. As it turned out, to prove convergence of the transformed truncated Poisson random variable to Normal distribution, we did not use its expectation and second moment.

## 8 Summary and conclusions

The work of the Elitism Levels Traverse Mechanism is based on identifying the ways to add at most one super-elite offspring to the population. Probabilities from these expressions serve as parameters in the birth-and-death MC. For the Royal Roads function with fitness plateaus we obtained an upper bound (worst-case analysis) in a very general form that is a function of the population  $\mu$  and recombination pool sizes  $\lambda$ , length of plateaus  $M$  and the number of fitness levels  $K$ . For a specific case of  $M = K = \sqrt{n}$  a sensible upper bound of  $O(\frac{\mu n \log^2 n}{\lambda})$  was derived that is better than  $O(n^2)$  bound previously known in literature.

Additionally, we have shown that, as the number of super-elite species increases (till  $\frac{\mu}{2}$ ), so does the probability of adding the super-elite offspring. If the run of the algorithm is broken down into two Phases: solution of the first active bin and the rest  $K - 1$  bins, the runtime of Phase 2 is only by a factor of  $\log n$  larger than Phase 1. This confirms the intuition that it takes much longer to solve the bin only when a relatively small number of 1-bits is available in the population (early phase) than when there are many species with a large number of 1-bits (later phase).

Perhaps for the first time in EA community we have attempted to derive the distribution of species in the population. Using the MC defined above, we have approximated the stationary distribution of super-elite species. We have shown that it converges to truncated Poisson distribution if the progress rate is small and Normal if it is large. This finding helps better understand the limiting structure of the population. It is certainly possible to do so for other subsets thereof, which will certainly contribute to a better understanding of how EAs work.

## References

- [CHS<sup>+</sup>09] Tianshi Chen, Jun He, Guangzhong Sun, Guoliang Chen, and Xin Yao. A New Approach for Analyzing Average Time Complexity of Population-Based Evolutionary Algorithms on Unimodal Problems. *IEEE Transactions on Systems, Man, and Cybernetics*, 39(5):1092–1106, 2009.
- [CTCY11] Tianshi Chen, Ke Tang, Guoliang Chen, and Xin Yao. A large population size can be unhelpful in evolutionary algorithms. In press. *Theoretical Computer Science*, 2011.



- [DJW10a] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Drift Analysis and Linear Functions Revisited. In *Genetic and Evolutionary Computing Conference (GECCO) 2010*, pages 1–8, 2010.
- [DJW10b] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Multiplicative Drift Analysis. In *Genetic and Evolutionary Computing Conference (GECCO) 2010*, pages 1449–1456, 2010.
- [DJW11] Benjamin Doerr, Daniel Johannsen, and Carola Winzen. Multiplicative Drift Analysis. arxiv:1101.0776v1. 2011.
- [GKS99] Josselin Garnier, Leila Kallel, and Marc Schoenauer. Rigorous hitting times for binary mutations. *Evolutionary Computation*, 7(1):45–68, 1999.
- [Haj82] Bruce Hajek. Hitting-Times and Occupation-Time Bounds Implied by Drift Analysis with Applications. *Advanced Applied Probability*, 14:502–525, 1982.
- [He10] Jun He. A Note on the First Hitting Time of  $(1 + \lambda)$  Evolutionary Algorithm for Linear Functions with Boolean Inputs. In *Conference on Evolutionary Computation (CEC)*, pages 1–6, 2010.
- [HY01] Jun He and Xin Yao. Drift analysis and average time complexity of evolutionary algorithms. *Artificial Intelligence*, 127(2001):57–85, 2001.
- [HY02] Jun He and Xin Yao. From an Individual to a Population: An Analysis of the First Hitting Time of Population-Based Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 6-5, October 2002:495–511, 2002.
- [HY03] Jun He and Xin Yao. Towards an analytic framework for analysing the computational time of evolutionary algorithm. *Artificial Intelligence*, 145(2003):59–97, 2003.
- [HY04] Jun He and Xin Yao. A study of drift analysis for estimating computation time of evolutionary algorithm. *Natural Computing*, 3(2004):21–35, 2004.
- [Jag08] Jens Jagerskupper. A Blend of Markov Chains and Drift Analysis. *Parallel Problem Solving from Nature X*, 5199/2008:41–51, 2008.
- [JDJW05] Thomas Jansen, Kenneth A. De Jong, and Ingo Wegener. On the Choice of the Offspring Population Size in Evolutionary Algorithm. *Evolutionary Computation*, 13(4):413–440, 2005.

- [JW01] Thomas Jansen and Ingo Wegener. Evolutionary Algorithms-How to Cope With Plateaus of Constant Fitness and When to Reject Strings of The Same Fitness. *IEEE Transactions on Evolutionary Computation*, 5(6):589–599, 2001.
- [MFH92] Melanie Mitchell, Stephanie Forrest, and John H. Holland. The Royal Road for Genetic Algorithms: Fitness Landscapes and GA Performance. *European Conference on Artificial Life*, 1:245–254, 1992.
- [Mit96] M. Mitchell. *Introduction to Genetic Algorithm*. Kluwer Academic Publishers, 1996.
- [Nas96] Ingemar Nasell. On the quasi-stationary distribution of the closed endemic SIS model. *Advances in Applied Probability*, 28(1996):895–932, 1996.
- [Nas99] Ingemar Nasell. On the quasi-stationary distribution of the stochastic logistic epidemic. *Mathematical Biosciences*, 156(1999):21–40, 1999.
- [NOW09] Frank Neumann, Pietro S. Oliveto, and Carsten Witt. Theoretical Fitness-Prpoportional Selection: Landscapes and Efficiency. In *Genetic and Evolutionary Computing Conference (GECCO) 2009*, pages 835–842, 2009.
- [OHY08] Pietro S. Oliveto, Jun He, and Xin Yao. Analysis of Population-based Evolutionary Algorithms for the Vertex Cover Problem. In *Congress of Evolutionary Computation(CEC)*, pages 1563–1570, 2008.
- [Rud98] Günter Rudolph. Finite Markov Chain results in evolutionary computation: A tour d’horizon. In *Fundamenta Informaticæ*, volume 1-22, 1998.
- [SW03] Tobias Storch and Ingo Wegener. Real Royal Road Function for Constant Population Size. In *Genetic and Evolutionary Computing Conference (GECCO) 2003*, pages 1406–1417, 2003.
- [TS12] Aram Ter-Sarkisov. Elitism Levels Traverse Mechanism For The Derivation of Upper Bounds on Unimodal Functions. In *WCCI 2012 IEEE World Congress on Computational Intelligence*, pages 2161–2168, 2012.
- [TSM11a] A. Ter-Sarkisov and S. Marsland. Convergence of a Recombination-Based Elitist Evolutionary Algorithm on the Royal Roads Test Function. In

*24th Australasian Joint Conference on Artificial Intelligence*, pages 361–371, 2011.

- [TSM11b] A. Ter-Sarkisov and S. Marsland. Convergence Properties of Two  $(\mu + \lambda)$  Evolutionary Algorithms on OneMax and Royal Roads Test Functions. In *International Conference on Evolutionary Computation Theorey and Applications (ECTA)*, pages 196–202, 2011.
- [TSMH10] A. Ter-Sarkisov, S. Marsland, and B. Holland. The k-Bit-Swap: A New Genetic Algorithm Operator. In *Genetic and Evolutionary Computing Conference (GECCO) 2010*, pages 815–816, 2010.
- [Wit06] Carsten Witt. An Analysis of the  $(\mu + 1)$ EA on Simple Pseudo-Boolean Functions. In *Genetic and Evolutionary Computing Conference (GECCO)*, pages 761–773, 2006.
- [Wit08] Carsten Witt. Population size versus runtime of a simple evolutionary algorithm. *Theoretical Computer Science*, 403(2008):104–120, 2008.
- [WJ07] Richard A. Watson and Thomas Jansen. A Building Block Royal Road Where Crossover is Provably Essential. In *Genetic and Evolutionary Computing Conference (GECCO) 2007*, pages 1452–1459, 2007.
- [WM04] John C. Wierman and David J. Marchette. Modeling computer virus prevalence witht a susceptible-infected-susceptible model with reintroduction. *Computational Statistics and Data Analysis*, 45(2004):3–23, 2004.

## A Runtime of Phase 1

Here we have some of the algebraic derivations from Section 6, most of which were done in MATLAB and MATLAB SYMBOLIC TOOLBOX.

The **first** expression of Phase 1 is

$$p_{\alpha^*, \alpha^*+1} = \binom{\frac{\lambda}{2}}{1} \left( \frac{2\alpha^*}{4\mu^2} \varphi_1 + \frac{2\alpha^*}{2\mu} \left( \frac{\mu - 1 - \alpha^*}{2\mu} \right) \varphi_2 + \frac{1}{4\mu^2} \varphi_3 + \frac{2}{2\mu} \left( \frac{\mu - 1 - \alpha^*}{2\mu} \right) \varphi_4 \right)$$

All species in the population have the same fitness, there are no  $\eta$  parents. The swap probabilities are defined as following:

$$\begin{aligned}\varphi_1 &= \frac{j}{n^2}(2K - 3 + j) \\ \varphi_2 &= \frac{j}{n^2}(K - 3 + j) \\ \varphi_3 &= \frac{2(M + 1 - j)(2K - 3 + j)}{n^2} \\ \varphi_4 &= \frac{(M + 1 - j)(K - 3 + j)}{n^2}\end{aligned}$$

The first expression is due to selecting a 1-bit in the active bin in  $\alpha^*$  and a 1 anywhere in the second parent  $\beta^*$ . This uses the pessimistic assumption that in all bins  $\beta^*$  parents have only two 1-bits. The second one is the same just the second parent is  $\gamma^*$ . The third is due to selecting a 0-bit in the next-best species and a 1 anywhere in the second parent (also next-best species, therefore multiplied by 2). The fourth swap probability is selecting a 0-bit in the  $\beta^*$  parent's active bin and 1 anywhere from  $\gamma^*$ . Therefore,

$$\begin{aligned}p_{\alpha^*, \alpha^*+1} &= \frac{\lambda}{8\mu^2}(2\alpha^*(\varphi_1 + (\mu - 1 - \alpha^*)\varphi_2) + \varphi_3 + (\mu - 1 - \alpha^*)\varphi_4) \\ &= \frac{\lambda}{8\mu^2 n^2}(j(2K - 3 + j) - \alpha^*j(K - 3 + j) + (\mu - 1)j(K - 3 + j)) \\ &\quad + (M + 1 - j)(2(2K - 3 + j) - 2\alpha^*(K - 3 + j) + 2(\mu - 1)(K - 3 + j)) \\ &= \frac{\lambda}{8\mu^2 n^2}(2j(\alpha^* - 1)(K + (K - 3 + j)(1 - \alpha^*) + 2j(\mu - 1)(K - 3 + j)(\alpha^* + 1))) \\ &\approx \frac{2\lambda j \alpha^*(K + (K - 3 + j)(\mu + 1 - \alpha^*))}{8\mu^2 n^2}\end{aligned}$$

The **first** expression of the expectation of Phase 1 is

$$S_1 = \sum_{j=3}^M \sum_{\alpha^*=1}^{\delta^* \alpha - 1} \frac{1}{p_{\alpha^*, \alpha^*+1}(j)} = \frac{8\mu^2 n^2}{\lambda} \sum_{j=3}^M \sum_{\alpha^*=1}^{\delta^* \alpha} \frac{1}{2j\alpha^*(K + (K - 3 + j)(\mu + 1 - \alpha^*))}$$

We continue by expanding the summand in partial fractions (w.r.t  $\alpha^*$ ):

$$\begin{aligned}S_1^* &= \frac{1}{2j\alpha^*(K + (K - 3 + j)(\mu + 1 - \alpha^*))} \\ &\approx \frac{1}{2j(K + K\mu + j\mu)} \left( \frac{1}{\alpha^*} + \frac{K + j}{K + K\mu + j\mu - \alpha^*(K + j)} \right)\end{aligned}$$

Sum over  $\alpha^*$  of the first fraction is of course  $\log(\delta^*\alpha)$ . The second fraction can be approximated in the following way (up to a constant):

$$\frac{K+j}{K+K\mu+j\mu-\alpha^*(K+j)} \approx \frac{1}{\mu-\alpha^*}$$

and the expression becomes

$$S_1^* \approx \frac{1}{2j(K+K\mu+j\mu)} \left( \log \frac{\delta^*\alpha\mu}{\mu-\delta^*\alpha} \right)$$

Again, expanding the expression in partial fractions w.r.t.  $j$ , we get

$$S_1^* \approx \log \left( \frac{\delta^*\alpha\mu}{\mu-\delta^*\alpha} \right) \left( \frac{1}{2(K+K\mu)j} - \frac{\mu}{2(K+K\mu)(K+K\mu+j\mu)} \right)$$

Summing the first term w.r.t.  $j$  we get  $H_M \approx \log(M+1)$ . The second term is at most  $O(1)$ . The upper bound for the **first** expression is (since  $\frac{\mu}{\mu+1} \approx 1$ )

$$\frac{4\mu n^2 \log(M+1) \log \frac{\delta^*\alpha\mu}{\mu-\delta^*\alpha}}{\lambda K}$$

The **second** expression in Phase 1 is

$$\frac{\delta^*\alpha}{p_{1,2}^{\delta^*\alpha}} \sum_{\alpha^*=2}^{\delta^*\alpha-1} q_{\alpha^*, \alpha^*-1}$$

We simplify the cumbersome term in the denominator in the second expression (assuming  $r_{1,1} \leq p_{1,2}$  and using Bernoulli inequality):

$$\frac{\delta^*\alpha}{p_{1,2}^{\delta^*\alpha}} = \frac{\delta^*\alpha}{(1-(1-p_{1,2}))^{\delta^*\alpha}} = \frac{\delta^*\alpha}{(1-r_{1,1})^{\delta^*\alpha}} \leq \frac{\delta^*\alpha}{1-\delta^*\alpha p_{1,2}}$$

The expression for  $p_{1,2}$  can be upper-bounded in the following way:

$$\begin{aligned}
\delta^* \alpha p_{1,2} &= \frac{\lambda \delta^* \alpha}{2} \left( \frac{2j(j-1+2(K-1))}{4\mu^2 n^2} + \frac{2(\mu-2)j(K-3+j)}{4\mu^2 n^2} \right. \\
&\quad + \frac{2(M+1-j)(2K-3+j)}{4\mu^2 n^2} \\
&\quad \left. + \frac{2(\mu-2)(M+1-j)(K-3+j)}{4\mu^2 n^2} \right) \\
&\leq \frac{\lambda \delta^* \alpha}{4\mu^2 n^2} \left( j(2K+j) + \mu j(2K+j) + (2M-j)(2K+j) \right. \\
&\quad \left. + \mu(2M-j)(2K+j) \right) \leq \frac{\lambda \delta^* \alpha}{4\mu^2 n^2} \left( 2\mu j(2K+j) + 2\mu(2M-j)(2K+j) \right) \\
&= \frac{\lambda \delta^* \alpha M(2K+j)}{\mu n^2}
\end{aligned}$$

Trivially  $j < M$ , so the whole expression becomes (since we have at most  $M-3$  bits to flip:

$$\sum_{j=1}^{M-3} \frac{\delta^* \alpha}{p_{1,2}^{\delta^* \alpha}} \leq \frac{\delta^* \alpha (M-4)}{1 - \frac{\lambda \delta^* \alpha M(2K+M)}{\mu n^2}}$$

Thus we have simplified the front term in the second expression. we now derive the rest of it. We pessimistically assume that if the super-elite species degrades, it is replaced by its offspring. The only type of pairs in Phase 1 that enable degrading of super-elite species is  $\langle \alpha^*, \gamma^* \rangle$ .

$$\begin{aligned}
P_{sel} &= \frac{\alpha^*(\mu-1-\alpha^*)}{4\mu^2} \\
q_{flip} &= \frac{j}{n} \cdot \frac{n-(j-2)-(K-1)}{n} = \frac{j(n-K+3-j)}{n^2} \\
q'_{flip} &= 1 - q_{flip} \\
a &= \frac{\lambda j(n-K+3-j)}{n^2}
\end{aligned}$$

$$\begin{aligned}
\sum_{\alpha^*=2}^{\delta^*\alpha-1} q_{\alpha^*,\alpha^*-1} &= \sum_{\alpha^*=2}^{\delta^*\alpha-1} \binom{\frac{\lambda}{2}}{1} P_{sel} (1 - P_{sel})^{(\frac{\lambda}{2}-1)} q_{flip} \leq \sum_{\alpha^*=2}^{\delta^*\alpha-1} \sum_{l=0}^{\frac{\lambda}{2}} \binom{\frac{\lambda}{2}}{l} P_{sel}^l (1 - P_{sel})^{(\frac{\lambda}{2}-l)} q_{flip}^l \binom{l}{l} \\
&= \sum_{\alpha^*=2}^{\delta^*\alpha-1} (1 - P_{sel} q'_{flip})^{\frac{\lambda}{2}} = \sum_{\alpha^*=2}^{\delta^*\alpha-1} \left(1 - \frac{\alpha^*(\mu-1-\alpha^*)}{4\mu^2} q'_{flip}\right)^{\frac{4\mu^2}{4\mu^2} \frac{\lambda}{2}} \\
&\leq \sum_{\alpha^*=2}^{\delta^*\alpha-1} e^{-\frac{\alpha^*(\mu-1-\alpha^*)\lambda q'_{flip}}{8\mu^2}} \approx \delta^*\alpha \int_0^1 e^{-\frac{x(\mu-1)(\mu-1-(\mu-1)x)\lambda q'_{flip}}{8\mu^2}} dx \\
&\leq \delta^*\alpha \int_0^1 e^{-\frac{\mu^2 x(1-x)\lambda q'_{flip}}{32\mu^2}} dx = \delta^*\alpha \int_0^1 e^{-\frac{x(1-x)\lambda q'_{flip}}{32}} dx \\
&= \frac{\delta^*\alpha 2\sqrt{2}F\left(\frac{\sqrt{a}}{8\sqrt{2}}\right)}{\sqrt{a}}
\end{aligned}$$

The second step is due to Binomial theorem  $(a+b)^n = \sum_k \binom{n}{k} a^k b^{n-k}$  and the third one is the definition of the exponential function:  $(1 - \frac{a}{n})^n \leq e^{-a}$ . Sum is approximated using Riemann sums. The last expression is Dawson's integral. For small  $a$ , which we would expect them to be given that if  $j \leq M$  then  $a = O(\frac{1}{n})$  Dawson's integral can be expanded in Taylor series:

$$\frac{\delta^*\alpha 2\sqrt{2}F\left(\frac{\sqrt{a}}{8\sqrt{2}}\right)}{\sqrt{a}} \approx \delta^*\alpha \left(1 - \frac{a}{192}\right) \leq \delta^*\alpha$$

Summing over  $j$  this becomes

$$\sum_{j=1}^{M-3} \sum_{\alpha^*=2}^{\delta^*\alpha-1} q_{\alpha^*,\alpha^*-1} = \delta^*\alpha(M-4)$$

The **second** expression is

$$S_2 = \frac{\delta^*\alpha(M-4)}{1 - \frac{\lambda\delta^*\alpha M(2K+M)}{\mu n^2}} \cdot \delta^*\alpha(M-4) = \frac{(\delta^*\alpha(M-4))^2}{1 - \frac{\lambda\delta^*\alpha M(2K+M)}{\mu n^2}} \quad (17)$$

Therefore the expected first hitting time of Phase 1 becomes

$$\mathbf{E}T_1 \leq \frac{4\mu n^2 \log(M+1) \log \frac{\delta^*\alpha\mu}{\mu-\delta^*\alpha}}{\lambda K} + \frac{(\delta^*\alpha(M-4))^2}{1 - \frac{\lambda\delta^*\alpha M(2K+M)}{\mu n^2}}$$

## B Runtime of Phase 2

We present here algebraic derivation for Phase 2. The **first** expression is very similar to that in Phase 1 up to a constant, so we only give the main results.

$$p_{\alpha^*, \alpha^*+1} \geq \left(\frac{\frac{\lambda}{2}}{1}\right) \left( \frac{2\alpha^*}{4\mu^2} \varphi_1 + \frac{4\alpha^*}{2\mu} \left( \frac{\mu-1-\alpha^*}{2\mu} \right) \varphi_2 + \frac{1}{4\mu^2} \varphi_3 + \frac{4}{2\mu} \left( \frac{\mu-1-\alpha^*}{2\mu} \right) \varphi_4 \right)$$

The swap probabilities are

$$\begin{aligned} \varphi_1 &= \frac{j(kM + 2K + j - 3)}{n^2} \\ \varphi_2 &= \frac{j(kM + K + j - 2)}{n^2} \\ \varphi_3 &= \frac{2(M - j + 1)(kM + 2K + j - 3)}{n^2} \\ \varphi_4 &= \frac{(M - j + 1)(kM + K + j - 2)}{n^2} \end{aligned}$$

So  $p_{\alpha^*, \alpha^*+1}$  can be transformed accordingly:

$$\begin{aligned} p_{\alpha^*, \alpha^*+1} &\geq \frac{\lambda}{8\mu^2} (2\alpha^* \varphi_1 + \varphi_3 + 4(\mu - 1 - \alpha^*)(\alpha^* \varphi_2 + \varphi_4)) \\ &= \frac{\lambda}{8\mu^2 n^2} (2(kM + 2K + j - 3)(M + 1 + j(\alpha^* - 1)) \\ &\quad + 4(\mu - 1 - \alpha^*)(kM + K + j - 2)(M + 1 + j(\alpha^* - 1))) \\ &= \frac{\lambda}{4\mu^2 n^2} (M + 1 + j(\alpha^* - 1))(K + (kM + k + j - 3)(2\mu + 3 - \alpha^*)) \end{aligned}$$

The **first** expression is

$$S_1 = \frac{4\mu^2 n^2}{\lambda} \sum_{k=1}^K \sum_{j=1}^M \sum_{\alpha^*=1}^{\delta^* \alpha - 1} \frac{1}{(M + 1 + j(\alpha^* - 1))(K + (kM + k + j - 3)(2\mu + 3 - \alpha^*))}$$

Now for the **second** expression:

$$S_2(n, K, M, \alpha) = \sum_k \sum_j \frac{\delta^* \alpha - 1}{p_{1,2}^{\delta^* \alpha}} \cdot \sum_{\alpha^*=2}^{\delta^* \alpha - 1} q_{\alpha^*, \alpha^*-1}$$



For the front term we use the same trivial upper bound, now for  $K$ :

$$\sum_{k=1}^{K-1} \frac{\delta^* \alpha (M-4)}{1 - \frac{\lambda \delta^* \alpha M (2K+M)}{\mu n^2}} = \frac{\delta^* \alpha (M-4)(K-2)}{1 - \frac{\lambda \delta^* \alpha M (2K+M)}{\mu n^2}}$$

There are two ways to lose a super-elite species now, when pairing it with either  $\gamma^*$  or  $\eta$ . The probability to select a 0-bit in  $\gamma^*$  parent is

$$\frac{n - (K-1) - k(M-1) - (j-2)}{n} = \frac{n - K + 3 - k(M-1) - j}{n}$$

for  $\eta$  it is

$$\frac{n - (K-1) - (k-1)(M-1)}{n}$$

which is larger than the one for  $\gamma^*$ . The probability to select a 1-bit in the active bin in  $\alpha^*$  parent is always  $\frac{j}{n}$ . For simplification we do not consider degrading of previously-improved bins, which would of course relegate the super-elite parent into  $\eta$  subset. We get:

$$\begin{aligned} S_2 &= \sum_k \sum_j \sum_{\alpha^*=2}^{\delta^* \alpha - 1} q_{\alpha^*, \alpha^* - 1} = \left(\frac{\frac{\lambda}{2}}{1}\right) P_{sel1} q_{swap1} + \left(\frac{\frac{\lambda}{2}}{1}\right) P_{sel2} q_{swap2} \\ &\leq \sum_k \sum_j \sum_{\alpha^*=2}^{\delta^* \alpha - 1} \lambda P_{sel1} q_{swap2} = \frac{\lambda}{\mu^2} \sum_{k=1}^{K-1} \frac{(n - K - 1 - (k-1)(M-1))}{n^2} \\ &\quad \cdot \sum_{j=3}^{M-1} \frac{j}{4} \sum_{\alpha^*=1}^{\delta^* \alpha - 1} \alpha^* (\mu - \alpha^*) \\ &= \frac{\lambda}{4\mu^2 n^2} \cdot \frac{M^2 - M - 6}{2} \cdot \frac{(3\mu - 2\delta^* \alpha - 1)\delta^* \alpha (\delta^* \alpha - 1)}{6} \cdot \frac{(K-2)(2n - K - KM + M - 3)}{2} \\ &\leq \frac{\lambda(M^2 - M - 6)(3\mu - 2\delta^* \alpha - 1)(\delta^* \alpha)^2(K-2)(2n - K - KM + M - 3)}{96\mu^2 n^2} \end{aligned}$$

The combined expression for Phase 2 is

$$\begin{aligned} \mathbf{E}T_2 &\leq \frac{8\mu n^2}{\lambda} \cdot \frac{4\delta^* \alpha}{2\delta^* \alpha K + M} \log \left( 3 + \frac{M^2}{n\delta^* \alpha} \right) + \frac{(M-4)(K-2)^2 \lambda (M^2 - M - 6)}{\left( 1 - \frac{\lambda \delta^* \alpha M (2K+M)}{\mu n^2} \right)} \\ &\quad \cdot \frac{(3\mu - 2\delta^* \alpha - 1)(\delta^* \alpha)^3 (2n - K - KM + M - 3)}{96\mu^2 n^2} \end{aligned}$$

## C Lower bounds on probabilities of $\beta$ species in Phase 2

In Section 6 for lower bounds on probabilities for pairs  $\langle \alpha^*, \eta \rangle$  and  $\langle \beta^*, \eta \rangle$  in Phase 2 derivation we used values for  $\langle \alpha^*, \gamma^* \rangle$  and  $\langle \beta^*, \gamma^* \rangle$ . We show here more rigorously for what values this bound is correct. We compare the following expressions:

$$P(\langle \alpha^*, \gamma^* \rangle) = \binom{\frac{\lambda}{2}}{1} \frac{\alpha^*(\mu - 1 - \alpha^*)}{4\mu^2} \varphi_2$$

$$P(\langle \alpha^*, \eta \rangle) = \binom{\frac{\lambda}{2}}{1} \frac{\alpha^*(\mu - \alpha)^2}{2\mu^3} \varphi_5$$

The swap probability  $\varphi_5 > \varphi_2$  since we need to select a 0 in  $\alpha^*$  active bin and a 0 in the second parent, and obviously there are more 0-bits in  $\eta$  than in  $\gamma^*$  and terms  $\binom{\frac{\lambda}{2}}{1}$  and  $\frac{\alpha^*}{2\mu}$  cancel out. What remains to show is for what  $\alpha \geq \alpha^*$

$$\frac{(\mu - \alpha)^2}{\mu^2} \geq \frac{(\mu - 1 - \alpha^*)}{2\mu}$$

This is quadratic inequality in  $\alpha$ , so we express it as (set  $\mu^2 + \mu + \mu\alpha^* = t$ ):

$$2\alpha^2 - 4\mu\alpha + t > 0$$

Since  $t > 0$ ,  $4\mu > 0$  and  $0 < 2 < \frac{16\mu^2}{4(\mu^2 + \mu + \mu\alpha^*)}$ , the only sensible solution is

$$1 < \alpha^* \leq \frac{4\mu - \sqrt{8\mu}}{4} \leq \frac{4\mu - \sqrt{16\mu^2 - 4 \cdot 2 \cdot (\mu^2 + \mu + \mu\alpha^*)}}{4} \approx .29\mu$$

So for  $1 < \alpha < .29\mu$  the inequality holds:

$$P(\langle \alpha^*, \eta \rangle) > P(\langle \alpha^*, \gamma^* \rangle)$$

At the same time,

$$P(\langle \beta^*, \gamma^* \rangle) = \binom{\frac{\lambda}{2}}{1} \frac{1 \cdot (\mu - 1 - \alpha^*)}{4\mu^2} \frac{(M - j + 1)(kM + j - 2)}{n^2}$$

$$P(\langle \beta^*, \eta \rangle) \geq \binom{\frac{\lambda}{2}}{1} \frac{1 \cdot (\mu - \alpha)^2}{2\mu^3} \frac{(M - j - 1)K}{n^2}$$

Since  $\varphi_6 \leq \varphi_4$  in this case (it is more likely to sample a 1-bit from  $\gamma^*$  rather than  $\eta$ , we need to account for these values too (note we take an extreme lower bound on the probability of sampling a 1-bit from  $\eta$ ,  $\frac{K}{n}$ ). Canceling out terms, we obtain, like before a quadratic inequality in  $\alpha$ :

$$6K\alpha^2 - 12\mu K\alpha + \mu(kM + j - 2)(1 + \alpha^* - \mu) + 6\mu^2 K > 0$$

solving this, as before, there exists only one sensible solution:

$$1 < \alpha < \mu - \frac{\mu}{\sqrt{6}} \sqrt{\frac{kM + j - 2}{K}} \leq \mu - \frac{\sqrt{\frac{\mu(kM + j - 2)(\mu - \alpha^* - 1)}{K}}}{\sqrt{6}}$$

by taking  $k = K, j = M - 1$  it seems that the worst upper bound on  $\alpha$  is

$$1 < \alpha < \mu(1 - .4 \cdot \sqrt{M + \frac{M}{K} - \frac{3}{K}})$$

which is valid for

$$1 - .4 \cdot \sqrt{M + \frac{M}{K} - \frac{3}{K}} > 0 \Leftrightarrow K \geq \frac{M - 3}{6.25 - M}$$

that is, only small  $M \leq 6$ , i.e. for problems with small bins similar or marginally more complicated than OneMax. Most  $\eta$  parents have more than  $K$  1-bits, but we neither make assumptions about their distribution, nor track the change in their number (unlike  $\alpha^*$ ). If, instead we assume that the number is  $\sqrt{M}K$ , the bounds become valid till  $M \leq 39$ . We will keep investigating ways of making this approach more general.